Detection of Multiple Sclerosis Lesions using Sparse Representations and Dictionary Learning

Hrishikesh Deshpande, Pierre Maurel, Christian Barillot

Detection of Multiple Sclerosis Lesions using Sparse Representations and Dictionary Learning

Hrishikesh Deshpande[1,2,3,4], Pierre Maurel[1,2,3,4], Christian Barillot[1,2,3,4]

1. University of Rennes 1, Faculty of Medecine, F-35043 Rennes, France
2. INSERM, U746, F-35042 Rennes, France
3. CNRS, IRISA, UMR 6074, F-35042 Rennes, France
4. Inria, VISAGES Project-team, F-35042 Rennes, France

Abstract. The manual delineation of Multiple Sclerosis (MS) lesions is a challenging task pertaining to the requirement of neurological experts and high intra- and inter-observer variability. It is also time consuming because large number of Magnetic Resonance (MR) image slices are needed to obtain 3-D information. Over the last years, various models combined with supervised and unsupervised classification methods have been proposed for segmentation of MS lesions using MR images. Recently, signal modeling using sparse representations (SR) has gained tremendous attention and is an area of active research. SR allows coding data as sparse linear combinations of the elements of over-complete dictionary and has led to interesting image recognition results. The dictionary used for sparse coding plays a key role in the classification process. In this work, we have proposed to learn class specific dictionaries and developed a new classification scheme, to automatically detect MS lesions in 3-D multi-channel MR images.

1 Introduction

Multiple sclerosis is a chronic, autoimmune disease of the central nervous system, characterized by structural damages of axons and their myelin sheathes. During progression of the disease, certain areas of brain develop MS lesions. The evolution of MS lesions is highly variable and is not fully known. MS is more common in North America and Europe and is more prevalent in young adult population, causing non-traumatic disabilities.

Magnetic Resonance Imaging (MRI) holds the capability of detecting abnormalities in 95% of the patients with MS and is the best paraclinical method for imaging MS [1]. These images are analyzed to find the number and spatial patterns of the lesions, appearance of new lesions and the total lesion load, which are key parameters in the current MS diagnostic setup. However, manual segmentation of MS lesions is a laborious and time consuming process and is prone to high intra- and inter-expert variability. Therefore, there is a need for fully automated MS lesion detection methods that can handle large variety of MR data and which can provide results that correlate well with expert analysis [2].

Over the past years, various approaches for semi-automatic and automatic segmentation of MS lesions have been proposed. In these methods, different image features, classification methods and models have been tried, but they usually suffer from high sensitivity to the imaging protocols and so usually require tedious parameter tuning or specific normalized protocols [3]. More recently, sparse representation has evolved as a model to represent an important variety of natural signals using few elements of an overcomplete dictionary. Many publications have demonstrated that sparse modeling can achieve state-of-the-art results in image processing applications such as denoising, texture segmentation and face recognition [4, 5]. In [5], given multiple images of individual subjects under varying expressions and illuminations, the images themselves were used as dictionary elements, for classification. Such a method uses dictionary learning to analyze image as a whole. Mairal et al [6] proposed to learn discriminative dictionaries better suited for local image descrimination tasks. In medical imaging, local image analysis is of prime importance and it could be interesting to see the performance of sparse representation and dictionary learning based classification methods in the context of disease detection. Some researchers have reported works on segmentation of endocardium and MS lesions using dictionary learning [7, 8]. Weiss et al. proposed an unsupervised approach for MS lesion segmentation, in which a dictionary learned using healthy brain tissue and lesion patches is used as basis for classification [7].

Our approach differs from this method in several ways. In [7], authors use only FLAIR MR images for analysis of clinical data. However, MS lesions appear in different intensity patterns in various MR sequences, which include T1 (T1-w MPRAGE) and T2-weighted, (T2-w) and Proton Density (PD). The complementary information in these MR images can further assist in classifying MS lesions. We build our analysis using above mentioned MR sequences. Our major contribution is however that we learn class specific dictionaries for healthy brain tissues and lesions that promote the sparse representation of healthy and lesion patches. The lesion patches are well adapted to its own class dictionary, as opposed to the other. Thus, we can use the reconstruction error derived from sparse decomposition of test patch on to these dictionaries for classification. In the dataset, the healthy class patches outnumber lesion patches and exhibit more variability. Thus we use different dictionary lengths for modelling individual class patches. In this manner, we take into consideration the data variability and class imbalance in healthy and lesion classes. Finally, supervised approach for detection results in omission of tuning of one parameter as mentioned in [7], making this method fully automatic. In the following sections, we describe our new approach and provide its evaluation using clinical images.

2 Methodology

As shown in Figure 1, we first preprocess MR images for noise removal and then extract the image patches of predefined size using brain mask. These patches are normalized and are divided into the training and test sets for healthy brain

tissue and lesion classes, with the help of manual segmentation images. Using training signals, we derive different classification approaches by either learning single dictionary or two separate dictionaries for both the classes. Finally, for a given test patch, the reconstruction error based classification method is developed, followed by voxel-wise classification and lesion detection. The following subsections briefly describe these steps.

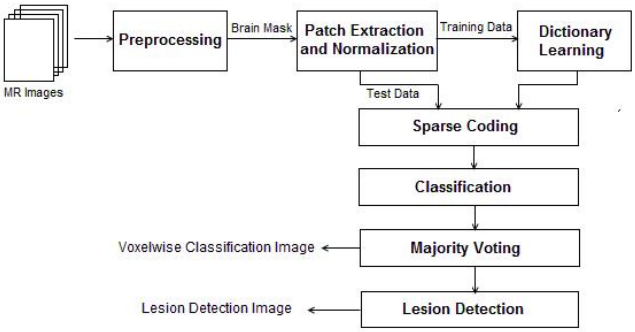


Fig. 1. Flowchart of MS Lesion Detection using Dictionary Learning

2.1 Patch Extraction and Training Set

We divide the intracranial MR volume into several 3-D patches and flatten them into one dimensional concatenated vectors representing intensities of T1-w MPRAGE, T2-w, PD and FLAIR images. Keeping the computational complexity of further analysis in mind, we extract a patch every M voxels in each direction. As described earlier, we develop supervised approach by labelling these patches as belonging to either healthy or lesion class. If, in a patch, the number of voxels manually labelled as lesions exceeds a threshold $T_L = 6mm^3$, it is included in a lesion set, or in healthy set otherwise. For every subject, we obtain around 1.5×10^6 patches for healthy and 10^3 to 10^5 patches for lesion class, depending on the lesion load for each patient. These patches are finally normalized to limit their individual norms below or equal to unity, as per constraint imposed by dictionary learning.

2.2 Sparse Representations and Dictionary Learning

Sparse representation of the data allows the decomposition of signal into linear combination of few basis elements in an overcomplete dictionary. Consider a signal $\mathbf{x} \in \mathbb{R}^N$ and an overcomplete dictionary $\mathbf{D} \in \mathbb{R}^{N \times K}$. The sparse coding problem can be stated as $min_{\mathbf{a}} \|\mathbf{a}\|_0$ s.t. $\mathbf{x} = \mathbf{Da}$ or $\|\mathbf{x} - \mathbf{Da}\|_2^2 \leq \varepsilon$, where $\|\mathbf{a}\|_0$ is l_0 norm of the sparse coefficient vector $\mathbf{a} \in \mathbb{R}^K$ and ε is error in representation.

Basis pursuit algorithm solves the convex approximation of the problem above by replacing l_0 norm with l_1 norm that also results in sparse solution [9]. Thus, the sparse coding problem can be given by

$$\min_{\mathbf{a}} \|\mathbf{x} - \mathbf{D}\mathbf{a}\|_2^2 + \lambda \|\mathbf{a}\|_1 , \tag{1}$$

where λ controls the trade-off between representation error and sparsity.

The fixed dictionaries like wavelets can be efficient if a background analytical model can be inferred. On the other hand, the dictionary learning from underlying data has produced exciting results with greater data adaptibility and has replaced the use of generic models. For a set of signals $\{\mathbf{x}_i\}_{i=1,..,m}$, the dictionary learning problem is to find $\mathbf{D}$ such that each signal can be represented by sparse linear combination of its atoms. This can be stated as the following optimization problem

$$\min_{\mathbf{D},\{\mathbf{a}_i\}_{i=1,..,m}} \sum_{i=1}^{m} \|\mathbf{x}_i - \mathbf{D}\mathbf{a}_i\|_2^2 + \lambda \|\mathbf{a}_i\|_1 . \tag{2}$$

The optimization is carried out as two-step process involving the sparse coding step with fixed $\mathbf{D}$ and the dictionary update step with fixed $\mathbf{a}$.

2.3 Classification

(a) Using Single Dictionary : In the context of MS lesion classification, the simplest idea, similar to [7], could be to use a single dictionary learned from healthy and lesion class patches. As the lesions are outliers with respect to the healthy brain intensities, the decomposition of lesion patch using this dictionary would result in higher representation error than that for the healthy tissue patch. For a given test patch, we calculate the sparse coefficients and reconstruction error, and assign it to the lesion class if this error is greater than chosen threshold. The threshold is selected by observing the histogram of the error map.

(b) Using Class Specific Dictionaries (Same Length) : Here, we learn class specific dictionaries D_1 and D_2 for healthy and lesion classes, respectively. Given a test patch $\mathbf{x} \in \mathbb{R}^N$, classification is performed in two steps: In the first step, sparse coefficients $\mathbf{a}_i$ are obtained using Eq (1) for each class i=1 (Healthy) and 2 (Lesion). The test patch is then assigned to class c such that

$$c = \underset{i}{\operatorname{argmin}} \|\mathbf{x} - \mathbf{D}_i\mathbf{a}_i\|_2^2 . \tag{3}$$

(c) Using Class Specific Dictionaries (Different Lengths) : The dictionaries learned using above mentioned approach does not take into account the data variability between two-classes. The size of the dictionary plays a major role in the data representation. For healthy class data with more variability and number of training samples than that for the lesion class, we allow larger dictionary length for healthy class data and study its effect on MS lesion classification.

2.4 Voxel-wise Classification and Lesion Detection

As already stated, there is some overlap between patches. However, to obtain voxel-wise classification, each voxel needs to be assigned to either of the classes. This is achieved using majority voting, in which, the voxel under consideration is classified as healthy or lesion, using majority votes of all patches which contain that voxel.

The voxelwise classification image is further processed to obtain the lesion based detection image. A lesion is said to be detected if $\frac{R_D \bigcap R_{GT}}{R_{GT}} \geq T_O$, where R_D and R_{GT} are respectively the candidate regions in the classification image and the ground truth, whereas T_O is the threshold indicating overlap between them as a fraction of ground truth lesion.

3 Dataset and Preprocessing

The proposed approach was validated on MRI volumes of 14 MS patients acquired by Verio 3T Siemens scanner. T1-w MPRAGE, T2-w, PD and FLAIR MR modalities were chosen for the experiment. The volume size for T1-w MPRAGE and FLAIR is $160 \times 256 \times 256$ and voxel size is $1 \times 1 \times 1$ mm^3, whereas for T2-w and PD scans, the volume size is $192 \times 256 \times 44$ and voxel size is $1 \times 1 \times 3$ mm^3. The manual segmentation images obtained from neurological experts are referred to as ground truth lesion masks.

For MR images of each patient, the imaging artifacts are corrected by denoising using non-local means and Intensity Inhomogenity Correction (IIH). The images so obtained are then registered with respect to T1-w MPRAGE volume and are processed further to extract the intra-cranial region.

4 Results and Discussions

We implemented our method using MATLAB and Python. The packages ANIMA and N3 ITK were used for denoising, registration and IIH correction, respectively [10–12]. We used the neuroimaging software Brain Extraction Tool (BET) for brain extraction [13]. For dictionary learning and sparse coding, we used SPArse Modeling Software (SPAMS) package [14].

We performed the experiments on 14 subjects using Leave-One-Subject-Out-Cross-Validation. Different parameters have been tested for the methods. It was found that image patch of size $5 \times 5 \times 5$, with a patch every 2 voxels in each direction, was optimal with respect to the classification efficiency. The dictionary length of 5000 and sparsity parameter $\lambda = 0.95$ were optimal selections for dictionary learning method. For voxel-wise classification method, we then recorded the number of voxels that belong to True Positives (TP), False Negatives (FN), False Positives (FP) or True Negatives (TN) and the classification methods were finally validated by calculating sensitivity= $\frac{TP}{TP+FN}$ and Positive Predictive Value (PPV) = $\frac{TP}{TP+FP}$.

In the first method, we studied the classification by learning single dictionary with the help of both healthy brain tissue and lesion patches. We chose sparse penalty factor $\lambda = 0.85$ in the sparse coding step and performed the classification for various threshold values on the histogram of error map, as explained previously. We then selected the threshold for which the best voxelwise classification results were obtained in terms of both sensitivity and PPV. It was observed that the method suffered with a very large number of false positive detections.

Next, we learned class specific dictionaries for healthy and lesion classes, each. We used dictionary lengths of 5000 for signal representation of each class. The mean sensitivity and PPV obtained using this approach were 91.5% and 7.5%. This method performs better than the previous method but still contains many false positives. The primary reason behind this can be the difference in the data variability of each class signals. The healthy class patches have more variability in terms of representation of white matter (WM), gray matter (GM) and cerebrospinal fluid (CSF), as compared to the variations in the representation of lesions. Hence, we adopted different dictionary lengths for representation of these classes. We used dictionary lengths of 5000 and 1000 respectively, for healthy and lesion classes. Table 1 summarizes the results of the voxelwise classification for the three methods described above.

Table 1. Voxel-wise classification results using: (a) Single Dictionary (SD), with 5000 atoms learned using healthy and lesion class data, (b) Class Specific Dictionaries with Same Lengths (CSD SL): 5000 atoms each and (c) Class Specific Dictionaries with Different Lengths (CSD DL): 5000 atoms for healthy class and 1000 atoms for lesion class. Sensitivity and Positive Predictive Value (PPV) (%) are given for each method.

Patient	(a) SD		(b) CSD SL		(c) CSD DL	
	Sens.	PPV	Sens.	PPV	Sens.	PPV
1	42	1	97	3	53	31
2	74	1	98	2	66	41
3	73	1	91	2	63	27
4	91	2	98	17	57	68
5	61	1	95	10	54	65
6	91	7	89	29	38	55
7	78	1	85	3	20	32
8	72	1	98	3	69	21
9	80	1	36	2	4	9
10	66	1	97	9	61	52
11	89	2	98	12	66	41
12	75	1	99	8	52	36
13	78	1	100	3	77	31
14	59	1	100	2	78	17
Average	**73.5**	**1.57**	**91.5**	**7.5**	**54.14**	**37.57**

It can be seen that using class specific dictionaries with the same dictionary length improves both sensitivity and PPV, as compared to the first method. But PPV in the second method is still low, indicating that there are still large number of false positives, which can explain higher sensitivity. Using different dictionary lenghts, as implemented in third method, drastically reduces the number of false positives, which can be seen by the significant increment in PPV, while keeping the sensitivity in the acceptable limit.

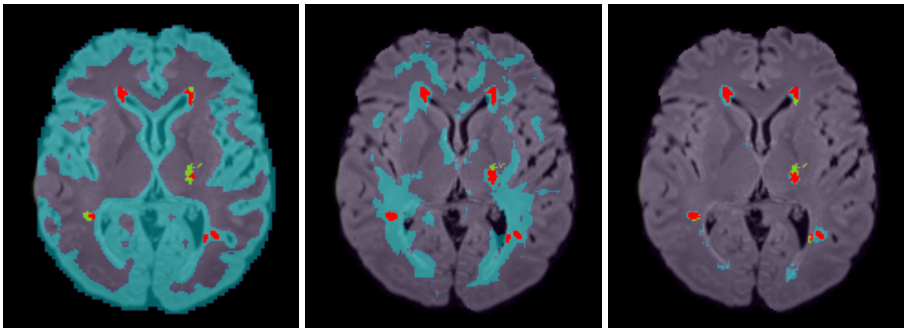
The mean PPV and sensitivity for lesion detection with class specific dictionaries of different lengths are shown in Table 2 for various overlap thresholds T_O. To be consistent with the threshold T_L incorporated in learning stage (Refer Section 2.1), we ignore very small lesions with volumes less than T_L. It can be seen that we detect 61% of the lesions with the overlap threshold of 1% . Moreover, in 49% of the lesions detected, at least 40% of the voxels are correctly classified by the method.

Table 2. Performance analysis for lesion detection using Class Specific Dictionaries with Different Lengths (CSD DL) for each class, with 5000 atoms for healthy class dictionary and 1000 atoms for lesion class dictionary.

	$T_O = 0.01$	$T_O = 0.1$	$T_O = 0.2$	$T_O = 0.3$	$T_O = 0.4$
PPV (%)	61.67	58.41	56.53	54.31	49.40
Sensitivity (%)	60.97	57.58	56.67	54.56	49.94

In Figure 2, we show the results for patient 8, for all the methods discussed above. The detection image is superimposed on FLAIR MR image. It can be observed that methods (a) and (b) have large number of false positives. We get the best classification results using class specific dictionaries with different dictionary lengths. But, in terms of voxelwise classification, there are still few false positives and true negatives around the actual lesion. This does not pose a major problem for lesion detection as long as significant portion of the actual lesion is being classified correctly. There are, however, some false positive lesion detections.

We are aware that we do not have a very large population for training. Hence we investigated the incorporation of longitudinal database into our analysis by considering MR sequences at 3 time points (M0, M3 and M6) for all the patients. As the lesions evolve over the course of time, it is fair to consider that each new dataset will enrich our learning model. Thus, we modified the training data, for each patient, in two ways: (1) Data at time-points M0 and M3, with 26 datasets and (2) Data at time-points M0, M3 and M6, with 39 datasets. However, the lesion detection experiments for the same test subjects, as in previous experiments, using class specific dictionaries with the lengths of 5000 and 1000 for healthy and lesion class respectively, did not show any significant improvement in the sensitivity and PPV. This suggests that the population for training the dictionaries earlier was sufficient and the dictionaries should be adapted to learn



(a) (b) (c)

Fig. 2. Classification results for Patient 8. For illustration purpose, one slice has been arbitrarily selected. True Positives are in red, False Positives are in cyan, False Negatives are in green. Methods (a), (b) and (c) are the same as in Table 1.

more specific structures viz. WM, GM and CSF versus lesions to help improve the detection.

5 Conclusion

In this paper, we have proposed a new supervised approach to automatically detect multiple sclerosis lesions using dictionary learning. We investigated the performance of three methods which either use one dictionary, treating lesions as outliers, or use class specific dictionaries for healthy and lesion classes, wherein the underlying data for each class is represented by the dictionary and sparse coefficients. We further studied the effect of using different dictionary lengths, allowing larger dictionaries to represent the complex data and concluded that such method minimizes the false positive detections in the classification.

Although the method using class specific dictionaries follows supervised approach, contrary to the single dictionary based classification method, which does not necessarily require training data, it is worth mentioning that the former method eliminates one parameter: threshold on error map. This crucial paramater is not easy to tune and could lead to worse classification results for small errors in the brain extraction procedure.

To further improve the results, it would be interesting not to learn only one dictionary for healthy brain tissues, but derive dictionaries more specific to WM, GM and CSF, in addition to the lesions. One could also study the role of sparse coefficients in addition to the reconstruction error, in the classification step.

References

1. Grossman, R., McGowan, J.: Perspectives on Multiple Sclerosis. AJNR (1998) 19(7):1251–1265

2. Polman, C., et al.: Diagnostic criteria for multiple sclerosis: 2010 Revisions to the McDonald criteria
3. García-Lorenzo, D., Francis, S., et al.: Review of automatic segmentation methods of multiple sclerosis white matter lesions on conventional magnetic resonance imaging. MedIA **17**(1) (2013) 1 – 18
4. Li, S., Fang, L., Yin, H.: An Efficient Dictionary Learning Algorithm and Its Application to 3-D Medical Image Denoising. Biomedical Engineering, IEEE Transactions on **59**(2) (Feb 2012) 417–427
5. Wright, J., Yang, A., Ganesh, A., Sastry, S., Ma, Y.: Robust Face Recognition via Sparse Representation. PAMI IEEE Transactions **31**(2) (Feb 2009) 210–227
6. Mairal, J., Bach, F., et al.: Discriminative learned dictionaries for local image analysis. In: IEEE CVPR. (June 2008) 1–8
7. Weiss, N., Rueckert, D., Rao, A.: Multiple Sclerosis Lesion Segmentation Using Dictionary Learning and Sparse Coding. In: MICCAI 2013. Volume 8149.
8. Rosas-Romero, R., Tagare, H.D.: Segmentation of endocardium in ultrasound images based on sparse representation over learned redundant dictionaries. Engineering Applications of Artificial Intelligence **29**(0) (2014) 201 – 210
9. Chen, S.S., Donoho, D.L., Michael, Saunders, A.: Atomic decomposition by basis pursuit. SIAM Journal on Scientific Computing **20** (1998) 33–61
10. Coupé, P., Yger, P., et al.: An optimized blockwise nonlocal means denoising filter for 3-D magnetic resonance images. IEEE TMI **27**(4) (April 2008) 425–41
11. William, W., Paul, V., et al.: Multi-modal volume registration by maximization of mutual information. MedIA **1**(1) (1996) 35 – 51
12. Tustison, N., Avants, B., , et al.: N4ITK: Improved N3 Bias Correction. IEEE TMI **29**(6) (June 2010) 1310–1320
13. Smith, S.: Fast robust automated brain extraction. In: HBM 17(3). (2002) 143–155
14. Mairal, J., Bach, F., et al.: Online Dictionary Learning for Sparse Coding. In: Proceedings of the 26th Annual International Conference on Machine Learning. ICML '09, New York, NY, USA, ACM (2009) 689–696